

Att bygga en språkbro mellan allmänhet och vårdpersonal

Språket i texter om hjärt-kärlsjukdomar

Dimitrios Kokkinakis, forskarassistent,
Maria Toporowska-Gronostaj, forskare samt
Sofie Johansson Kokkinakis, fil. dr i språkvetenskaplig
databehandling, samtliga verksamma vid institutionen
för svenska språket

Patienters delaktighet och inflytande på sin egen vård är en förutsättning för att framtidens hälso- och sjukvård ska uppskattas av medborgarna och dessutom vara kostnadseffektiv.

Konsumtionen av hälso- och sjukvård har ökat på senare år i Sverige. Det har medfört ett ökat intresse för medicin och vård. Människan är inte längre en passiv hälsokonsument utan förväntas vara en medveten och aktivt medverkande part i frisk- och sjukvården. Detta ställer allt större krav på patienters och deras närståendes förmåga att i detalj förstå all den hälsorelaterade information som de dagligen exponeras för. Denna information kan vara mer eller mindre lätt att förstå dels beroende på dess utformning, dels på mottagarens förkunskaper.

För att underlätta kommunikationen mellan hälsotagare och hälsogivare satsas det i utlandet och i Sverige på nätbaserad patientinformation. Det innebär att organisationer som läkemedelsproducenter, Socialstyrelsen och media via sina informationskanaler tillhandahåller mängder av information som riktas speciellt till hälsokonsumenterna. Information förmedlas främst språkligt men också visuellt genom bilder eller filmer på webben. Sådana resurser är mycket värdefulla och ger grundläggande kunskaper om olika åkommor och behandlingsformer. I några fall åtföljs informationen

av en ordlista där de viktigaste termerna finns förklarade. Sådana ordlistor fungerar som "första hjälpen" för att förstå informationen. Sökningar på nätet efter mer täckande medicinska ordböcker med integrerad utförlig språklig och medicinsk information för lekmän visar dock att sådana svenska resurser lyser med sin frånvaro. Av den anledningen kan det vara berättigat att reflektera över språket i medicinska texter för att i nästa steg kunna bygga upp textbaserade relevanta medicinska ordböcker med inslag av språklig och encyklopedisk medicinsk information för lekmän. Därför är ett av syftena med vår forskning att bygga en bro i form av en dynamisk elektronisk medicinsk ordbok som ska avspegla det språk som används på webben för medicinsk kommunikation.

För att kartlägga det medicinska språket har vi gjort ett antal kvantitativa och kvalitativa undersökningar som utgår ifrån analys av ordförrådet i två typer av medicinska textsamlingar inom området *hjärt-kärlsjukdomar*. Den ena textsamlingen riktar sig till hälsotagare och den andra till hälsogivare. Resultaten av våra undersökningar avslöjar en hel del karakteristiska drag och tendenser i de medicinska texterna. De frågor som vi försöker besvara i vår analys är:

- Vilka gemensamma specifika språkliga kännetecken finns i dessa texter?
- Vilka begrepp används av medicinska experter i deras interaktion med å ena sidan medicinskt kunniga och å andra sidan lekmän?
- Vilka benämningar finns för att kommunicera dessa begrepp?
- Vilka typiska kännetecken skiljer dessa två typer av texter åt?

Urval ur ett medicinskt empiriskt textmaterial

En textkorpus är en samling skrivna elektroniska texter som bearbetas för att underlätta datorbaserade studier av språket. En sådan korpus, som kartlägger sammansättningen av det svenska medicinska skrivna språket, har under de senaste två åren byggts upp och bearbetats på Institutionen för svenska språket. Korpusen utgör ett nödvändigt underlag för konstruktion av t. ex. term- och fraslistor, för studier av syntaxen och semantiken, men också för språkforskning av teoretiskt intresse. Korpusen är hämtad från webben och består av hela, sammanhängande texter, vilka representerar olika medicinska texttyper, som t. ex. beskrivning av sjukdomar, läkemedel, nya forskningsrön, föreskrifter osv. Det totala empiriska materialet

<i>Lekmannakorpus</i>	<i>Facktextkorpus</i>
När man får en blodpropp, så kallad trombos , leverar sig blodet och bildar en propp inuti ett blodkärl. Proppen gör det svårare för blodet att passera genom blodkärlet. Blodproppar är vanligast i benen, och börjar ofta i vaderna. Proppar kan någon gång bildas också på andra ställen i kroppen, till exempel i armarna, men det är inte alls lika vanligt. (Ur <i>Sjukvårdsrådgivningen</i> ; http://www.sjukvardsradgivningen.se)	Förhöjda nivåer av D-dimer i blodet är en indikator på fibrinolytisk brytning in vivo, och analysen har väckt stora förhoppningar vid diagnostik av venös tromboembolism. Den primära frågan är om analysen kan vara användbar för att utesluta djup ventrombos och lungembolism vid en akutmottagning. (Ur <i>Socialstyrelsens riktlinjer för vård av blodpropp/venös tromboembolism 2004</i> ; http://www.socialstyrelsen.se)

består av 11 miljoner ord. Texterna är strukturellt och i viss mån även innehållsligt kodade i enlighet med internationella riktlinjer med hjälp av textanalysverktyg (t. ex. ordklassbestämning). Eftersom den aktuella studien fokuserar på och jämför texter inom området hjärt-kärlsjukdomar, har vi från korpusen plockat texter som handlar om ämnet och vars huvudrubrik innehåller minst ett av tio utvalda nyckelord som kännetecknar området, inklusive sammansättningar med dessa ord. Dessa tio ord är: *hjärt-, propp, stroke, hemostas, koagulantia, stenos, trombos* och substantiverna/läkemedlen *fragmin, heparin* och *waran*. Beroende på källa delas texterna in i antingen en *facktextkorpus*, *FK*, eller en *lekmannakorpus*, *LK*. *FK* består av texter avsedda för läkare, t. ex. från *Läkartidningen* medan *LK* är texter avsedda för patienter, t. ex. från *NetDoktor.se*. Texterna består av drygt 80 000 ord vardera.

Läsbarhet och frekvensband – texternas svårighetsgrad

För att kunna få en första överblick över sammansättningen av texterna har vi använt ett läsbarhetsmått och en placering av ord i så kallade frekvensband. Ett frekvensband är en sorterad lista över alla ord i en text med de mest frekventa orden först. Nationalencyklopedin beskriver begreppet *läsbarhet* som 'ett mått på en texts svårighetsgrad'. Det finns flera olika sätt att beräkna detta, oftast med så kallade läsbarhetsindex (LIX). I Sverige är sLIX, ett svenskt LIX, det mest använda. sLIX tar hänsyn till två egen-

skaper hos texten; det *genomsnittliga antalet ord per mening* samt den procentuella *andelen långa ord* (ord längre än sex bokstäver). Genom att lägga samman dessa två värden får man en siffra som anger textens allmänna svårighetsgrad. Erfarenhetsmässigt vet man att texter med sLIX under 25 uppfattas som mycket lätta (t. ex. barnböcker), texter med sLIX mellan 25–35 som lätta (t. ex. lättlästa böcker). Texter med sLIX mellan 35–45 som medelsvåra (t. ex. dagspress), texter med sLIX mellan 45–55 som svåra (t. ex. officiella skrivelser) och texter med sLIX över 55 som mycket svåra (t. ex. facklitteratur). I enighet med vad vi förväntade oss i form av skillnad mellan de två texttyperna, visade resultatet att sLIX-värdet för FK var 48 och för LK var 38. Det bör betonas här att LIX är ett allmänt mått, som inte kan ange hur en bestämd person uppfattar en text. Det tar inte heller hänsyn till typografisk information som t. ex. bilder och layout.

Genom att använda frekvensband har vi beräknat överlappningen, och därmed frekvensfördelningen i ordförrådet mellan de två korpusarna i förhållande till de 10 000 mest frekventa orden i hela den medicinska korpusen. Det har visat sig att orden i den första gruppen som innehåller de 1 000 mest frekventa orden motsvarar då 73 procent av alla ord i FK och 80 procent i LK. Alltså utgör dessa ord, som hör till ett grundläggande ordförråd, den största delen av textmaterialet. Kännetecknande för det grundläggande ordförrådet är att:

- Alla ordklasser finns representerade
- De flesta ord är icke-sammansatta
- Majoriteten utgörs av svenska ordformer
- Allmän vokabulär dominerar, men pga. av texternas natur förekommer också ord som syftar till kroppsdelar, sjukdomar och medicinering.

I de mellanliggande frekvensbanden kunde vi observera att:

- Det i båda texterna endast finns en del ordklasser representerade (främst substantiv, verb och adjektiv). Ord hörande till de övriga ordklasserna finns bland de 1 000 mest frekventa orden.
- Antalet sammansättningar ökar snabbt i stigande frekvensband i båda texterna, men att antalet unika sammansättningsformer som t. ex. *trombosprofylaxen* och *trombosrecidiv* är mer framträdande i FK. Med andra ord finns det mindre upprepning och större variation i FK.

- Förekomsten av medicinska begrepp som t. ex. *djup ventrombos, cerebral trombos, Adam-Stokes syndrom* är mycket högre i FK.
- Latinska och grekiska ord är mer frekventa i FK, t. ex. *embolus, embolilektomi* och *sinus coronarius*.
- Nominaliseringar dominerar i FK. Det finns fem gånger fler substantiv än verb i t. ex. frekvensband 8 (det vill säga de ord som ligger mellan de 7 000 och 8 000 mest frekventa) i FK jämfört med fyra gånger fler i motsvarande frekvensband för LK.
- Antalet förkortningar och akronymer ökar stadigt. Antalet akronymer är fyra gånger större i FK än i LK.

Den sista och mest intressanta gruppen är frekvensband 10, det vill säga den grupp som innehåller de ord som inte är gemensamma för de båda texterna. Undersökningsresultatet visar att ord som finns i den gruppen och som är unika för respektive text är dubbelt så många i FK jämfört med de i LK. Denna grupp innehåller ofta mycket specialiserade ämnestypiska och ämnesspecifika termer.

Sammanfattningsvis kan man konstatera att de ordformer som kräver särskild uppmärksamhet vid design av ett medicinskt lexikon enligt den fördelning som nyss presenterades är:

- Unika eller icke-frekventa akronymer/förkortningar, t. ex. *ESCS, TMR, bitr.*
- Sammansättningar med eller utan akronymer, t. ex. *ICD-grupp, NSAID-preparat, angiotensinreceptorblockerare, vänsterkammarrekonstruktion.*
- Medicinska icke sammansatta ordformer, t. ex. *tromb, venös, klaff.*
- Frekventa lånord, t. ex. *stroke, angina pectoris, angiogram.*
- Felstavningar (både medicinska och allmänna ord), t. ex. *massspektrometri, kolecystekomi.*

Ordnivåundersökningar

Genom att beräkna olika statistiska mått på enskilda ord i de två texttyperna kan man få en bra uppfattning om ordförrådets komplexitet och därmed finna tydliga kännetecken som skiljer fackspråkliga texter från icke-fackspråkliga.

- LK bestod av färre antal ordformer men dessa upprepas oftare (9 gånger) än i FK (7 gånger).
- Den genomsnittliga längden på ordformer i FK var längre än i LK (9,5 gentemot 8,6); långa ordformer är utmärkande för facktexter.
- Sammansättningsformerna utgjorde 12 procent (36 procent unika) av ordformerna i FK jämfört med 10 procent (31 procent unika) i LK.
- Lexikal variation är ett mått på hur många likadana ord textförfattaren använt. Ju lägre värde desto fler lika ord i texten. Därmed ser man om textförfattaren upprepat sig sällan eller ofta. Distributionsanalysen visar att värdet i FK var 7 jämfört med 9 i LK.
- I FK fanns det en överrepresentation av akronymformer i sammansättningar, som t. ex. *TNF-alfa*, *WPW-syndrom* och *BNP-test*. Det finns också stora skillnader i antal ”enkla” akronymer, som t. ex. *ASA*.

Adjektiviska attribut

Adjektiviska och participiella attribut förstärker eller kompletterar innebörden hos många substantiviska huvudord. En viktig aspekt som kan kopplas till området fraser är så kallade *kollokationer*, dvs. ordkombinationer bestående av två eller flera ord som vanemässigt förekommer tillsammans i det talade eller skrivna språket. Vid en frekvensjämförelse mellan de två texterna fann vi att det fanns fler adjektiv i FK (8 400 olika böjningsformer – 1 700 grundformer) än i LK (6 800 olika böjningsformer – 1 200 grundformer). De mest statistiskt signifikanta attributen i FK var: *klinisk*, *randomiserad*, *diagnostisk*, *prediktiv*, *asymtomatisk*, *malign*, *kateterburen*, *neuropsykiatrisk*, *ischemisk* och *epidemiologisk*, medan de i LK var: *kallad*, *hämtad*, *mycken*, *vanlig*, *läckande*, *blodförtunnande*, *lätt*, *oregelbunden*, *bra* och *svår*. Listorna visar tydliga tecken på semantiska olikheter. I FK har attributen en tydlig fackspråksprofil, medan de i LK är mer generellt deskriptiva. Bristen på överlapp mellan listorna visar också olika kollokationspreferenser, ett viktigt element vid skapande eller förbättring av lekmanorienterade ordböcker och inte minst uppbyggande och förbättring av termlistor. Allt detta är också till stöd för att kunna bygga, förbättra eller utvärdera metoder och datorsystem som analyserar och försöker ”förstå” det medicinska språket.

Syntax och semantik

Inte bara frekvensanalys av ordförrådet utan också av syntaxen kan visa tecken på skillnader mellan texterna. En hög andel nominalfraser, bisatser och passiva konstruktioner anses vara karakteristiskt för fackspråkstexter. Texterna analyserades syntaktiskt med hjälp av datorprogram. Antal nominalfraser där huvudordet var ett medicinskt begrepp, underordnande satser och antal aktiva huvudsatser var högre i LK. Skillnaden mellan passiva konstruktioner var liten, 1 200 i FK jämfört med 1 100 i LK. Däremot var ett stort antal nominal- och prepositionsfraser dominerande i FK (17 400 och 9 800 gentemot 13 800 och 7 900 i LK). Intressant var den stora förekomsten av underordnade satser med *om man* i LK (250 förekomster jämfört med 20 i FK). Detta kännetecknar det förklarande i LK-korpusens natur.

För den semantiska analysen använde vi MeSH (Medical Subject Headings), den kontrollerade vokabulär som används vid U.S. National Library of Medicine, där de engelska termerna kompletterats med svenska översättningar gjorda av personal vid Karolinska institutets bibliotek. I FK fanns 4 620 fullständiga MeSH-matchningar (t. ex. "<artär>", och 409 partiella MeSH-matchningar (t. ex. "sub<araknoid>"), medan det i LK fanns 6 144 fullständiga och 277 partiella matchningar. Trots att antalet fullständiga matchningar i LK var mycket högre, visar antalet *unika* matchningar att det fanns högre antal *olika* termer i FK och högre andel upprepning i LK. Följaktligen är terminologin mer varierad i FK.

Avslutning

Vad kom vi då fram till i denna undersökning? Enligt fackspråksforskare finns det tydliga lingvistiska och lexikaliska kännetecken som skiljer fackspråkliga texter från andra texter. Dit hör bland annat förekomsten av nominala fraser, andelen facktermer, morfologiska, syntaktiska och semantiska egenskaper, menings- och ordlängd m. m. Vår undersökning av medicinska texter avsedda för vårdgivare respektive lekmän har bekräftat att de allra flesta generella språkliga tendenser som observerats i andra fackområden med några få undantag också stämmer på de undersökta medicinska texterna. Vår analys har visat att LK-texter verkar vara mer lätlästa än motsvarande FK-texter, men vi har inte empiriskt undersökt graden

av textförståelse i t. ex. en testläsargrupp. När det gäller ordförrådet visar det sig att endast en femtedel av orden i dessa texter som kan vålla problem, resten hör till allmänspråket. Men denna femtedel innehåller ganska många termer som är mycket relevanta i sammanhanget, eftersom de betecknar sjukdomar, symptom, anatomi, undersöknings- och behandlingsformer. För att underlätta förståelse av medicinska termer för den breda allmänheten skapas ofta förenklade benämningar på grekiska eller latinska termer, vilket bidrar till en ökning av synonyma termer som utgör länkar mellan fackspråket och det allmänna språket. Ett exempel är fackspråkets *trombos* som står för *blodpropp*, eller i vardagligt tal för *propp*. Vi har nu gjort en kartläggning av två typer av medicinska texter. Vi har identifierat vilka ord och termer som bör finnas med i ett webbaserat lexikon. Därför hoppas vi att detta kan förverkligas inom en snar framtid.